

Uncovering Phase Change Memory Energy Limits by Sub-Nanosecond Probing of Power Dissipation Dynamics

Keren Stern, Nicolás Wainstein, Yair Keller, Christopher M. Neumann, Eric Pop, Shahar Kvatinsky, and Eilam Yalon*

Phase change memory (PCM) is one of the leading candidates for neuromorphic hardware and has recently matured as a storage class memory. Yet, energy and power consumption remain key challenges for this technology because part of the PCM device must be self-heated to its melting temperature during reset. Here, it is shown that this reset energy can be reduced by nearly two orders of magnitude by minimizing the pulse width. A high-speed measurement setup is utilized to probe the energy consumption in PCM cells with varying pulse width (0.3–40 nanoseconds) and uncover the power dissipation dynamics. A key finding is that the switching power (P) remains unchanged for pulses wider than a short thermal time constant of the PCM ($\tau_{th} < 1$ ns in 50 nm diameter device), resulting in a decrease of energy ($E = P \times \tau$) as the pulse width τ is reduced in that range. Thermal confinement during short pulses is achieved by limiting the heat diffusion time. The improved programming scheme reduces reset energy density below 0.1 nJ μm^{-2} , over an order of magnitude lower than state-of-the-art PCM, potentially changing the roadmap of future data storage technology and paving the way toward energy-efficient neuromorphic hardware.

ratio ($>10^3$) between the amorphous and crystalline phase of chalcogenides glasses, mainly compounds of the ternary $\text{Ge}_x\text{Sb}_y\text{Te}_z$. The phase transition is triggered by Ovonic threshold switching,^[1,11] where the resistance of the chalcogenide is reduced above a certain threshold voltage, enabling high current densities, and self-heating for crystallization.

The memory cell structure consists of a phase change material, such as, $\text{Ge}_2\text{Sb}_2\text{Te}_5$ (GST), sandwiched between two electrodes. Electrical current crowding is achieved by either a small-area heater bottom electrode (BE) in a mushroom-type cell, or a narrow pore in the surrounding dielectric of the PCM layer in a so-called confined cell.^[2] To crystallize the PCM (i.e., set) to its low resistance state (LRS), a medium-amplitude voltage or current pulse is applied, which heats the material above its crystallization temperature. The set pulse must be sufficiently long to allow for the material to crystallize. To amor-

phize the cell (i.e., reset) to its high resistance state (HRS), a large-amplitude electrical pulse is applied for a short time (or with a short fall time) to melt the material and rapidly quench it below the crystallization temperature.^[2,12,13]

Because the phase transition is thermally induced, device-level heat management is crucial. The energy and power consumption are of great concern, and their fundamental limits are yet to be fully understood.^[14] Particularly, the reset process includes heating the phase change material above its melting temperature (typically $T_m > 600$ °C) which requires very high power density, on the order of ≈ 10 MW cm^{-2} (100 mW μm^{-2}). Reports on the values of reset power (during the pulse) and energy are scarce because of the need to capture the current and voltage during a short transient (typically less than 100 ns), but the reported energies per area are in the range of ≈ 1 J cm^{-2} (10 nJ μm^{-2}).^[15,16] Previous research efforts toward improving energy-efficiency (heating energy per device area) of the melt-quench process in PCM focused on thermal engineering of materials,^[17,18] interfaces,^[19,20] and device structure.^[21,22] However, the range of thermal properties of materials and their interfaces is limited.^[23]

In this article, we probe the power dissipation dynamics of PCM reset with sub-nanosecond (ns) resolution and show that the energy efficiency improves with the reduction of

1. Introduction

Phase change memory (PCM) stands as a promising candidate for storage,^[1,2] neuromorphic computing,^[3] radio-frequency applications,^[4,5] and in-memory computing,^[6–8] thanks to its nonvolatility, long retention, high endurance, short switching time, and compatibility with the back-end of line of standard silicon processing.^[9] Intel's Optane memory, a PCM-based memory aimed at storage class memory applications,^[10] is an example of the maturity of this technology. PCM is a two-terminal nonvolatile memory that relies on the high resistivity

K. Stern, N. Wainstein, Y. Keller, Prof. S. Kvatinsky, Prof. E. Yalon
Department of Electrical Engineering
Technion – Israel Institute of Technology
Haifa 32000, Israel
E-mail: eilamy@technion.ac.il

Dr. C. M. Neumann, Prof. E. Pop
Department of Electrical Engineering
Stanford University
Stanford, CA 94305, USA

 The ORCID identification number(s) for the author(s) of this article can be found under <https://doi.org/10.1002/aelm.202100217>.

DOI: 10.1002/aelm.202100217

programming pulse width, for example, the energy consumption can be reduced by approximately 100× as the pulse-width (PW) is reduced from 40 to 0.4 ns. Although sub-ns reset in PCM was already shown in several studies,^[24–28] the power and energy consumption during the sub-ns pulse application were not characterized to date. Measuring the current and voltage across the PCM during a sub-ns pulse is challenging because of parasitic capacitance and signal reflections at high frequencies (>1 GHz) in conventional on-wafer probing.

To circumvent this challenge, here we utilize a high-speed measurement setup, enabling us to directly probe the power and energy consumption in sub-ns reset pulses by measuring the current and voltage across the PCM during the applied pulse. We show that the reset power remains unchanged while reducing the PW below 5 ns for devices with via size in the range 50–200 nm. The reset power starts to increase only for PW shorter than the PCM thermal time constant (τ_{th}), namely $\approx 2\text{--}3$ ns and sub-1 ns for 200 and 50 nm nominal via size, respectively. As a result, the optimal energy-power-efficiency point can be traced. Although the power consumption increases as the PW is reduced below τ_{th} , the energy consumption continues decreasing with reduced PW, within the measurement limit of our experimental setup (0.3 ns).

We also perform finite-element method (FEM) simulations which predict the minimum reset power and energy, in good agreement with our experimental results. In addition to the characterization of power and energy consumption, our high-speed setup also allows us to study the thermal dynamics of the PCM. A drop in the transient resistance of the device during reset can be associated with melting of a critical volume of the GST, for example, $T > T_m$ (where T_m is the temperature required to melt the GST). Measurements of the transient resistance therefore allow us to determine the thermal time constant of the device, which is correlated with the optimal pulse for power-efficiency. Finally, reset pulses with the shortest PW of 0.3 ns (limited by instrumentation) achieve record-low energy consumption per area, namely ≈ 0.2 pJ for confined PCM cells with ≈ 50 nm via diameter or ≈ 0.1 nJ μm^{-2} , nearly two orders of magnitude below typical PCM reset energy density.^[15]

2. Results and Discussion

2.1. High-Speed Measurement Setup and Device Structure

The measurement setup (Figure 1a) consists of a fast pulse generator (PG) connected in series with the PCM device and a high-speed oscilloscope (see Section 4). To support sub-nanosecond pulses, ground-signal (GS) transmission lines with characteristic impedance $Z_0 = 50 \Omega$ were used to connect the device with the measurement equipment. Transmission lines reduce pulse broadening, signal reflections, and possible ringing due to parasitic capacitances. The output impedance of the PG, as well as, the input impedance of the scope are also set to 50Ω .

In this work, we focus on confined PCM cells as shown in Figure 1b,c, but the results are not limited to a specific PCM structure. The devices were fabricated as follows. First, tungsten (W) was evaporated, patterned, and etched to form the bottom electrode (BE). Next, SiO_x was deposited using plasma-

enhanced chemical vapor deposition (PECVD) and the confined vias were patterned using e-beam lithography. Sputtering and lift-off were used to pattern the GST layer with TiN capping, as well as, the final TiN/Pt top electrode (TE) and contact pads. More fabrication details are given in Section 4. The confined vias have different diameters, nominally: 50, 75, 100, and 200 nm, and the contact area of each set of devices was imaged using scanning electron microscopy (SEM) in order to measure their actual size, more details can be found in Figure S1 and Table S1, Supporting Information.

The devices are initially in the LRS (i.e., set, crystalline) of few k Ω (Figure 1d). We measure the PCM voltage and current during pulse application (Figure 1e) to obtain the transient resistance, power, and energy consumption. Following a successful reset pulse, the device is switched to the HRS (i.e., amorphous) of hundreds of k Ω to few M Ω (Figure 1f). Representative device endurance with sub-ns reset pulses is shown in Figure 1g preserving a resistance ratio of ≈ 100 within the first 1500 switching cycles, and more endurance data can be found in Figure S2, Supporting Information. The measurements outlined below are performed within the first 1000 cycles of each device, so no noticeable degradation is expected.

Our setup allows the measurement of transient PCM voltage $V_{\text{PCM}}(t)$ and current $I(t)$, hence transient power $P(t)$ and resistance $R_{\text{trans}}(t)$ are obtained directly, and the energy can be extracted by integrating the power over time. We define $V_{\text{PG}}(t)$ as the voltage applied by the pulse generator onto the 50Ω output resistance, and $V_{\text{scope}}(t)$ as the voltage measured by the scope across the $Z_0 = 50 \Omega$ load. The current is

$$I(t) = V_{\text{scope}}(t)/Z_0 \quad (1)$$

and the voltage across the PCM is

$$V_{\text{PCM}} = V_{\text{PG}} - I(t) \cdot (Z_0 + R_s) - V_{\text{scope}}(t) \quad (2)$$

where $R_s \approx 400 \Omega$ is the serial resistance of the lead, a thin W line connecting the BE to the pad, measured separately for each via size in devices without GST.

2.2. Record-Low Reset Energy Density in Phase Change Memory by Sub-ns Programming

Typically, PCM devices with 10 to 100 nm via diameter require 10 to 100 pJ for reset,^[15,16] whereas ≈ 0.1 pJ have been reported for PCM with carbon nanotube (CNT) electrodes thanks to the smaller volume to melt.^[29] Figure 2a,b show a comparison between the reset energy for different PCM structures versus effective contact diameter. The purple markers represent results from the literature,^[13–18,22,30–35] showing the scaling of reset energy with the effective diameter of the cell because less energy is needed to heat a smaller volume. Scaling cell dimensions also reduces the set energy. Low reset current and energy were achieved in doped GST, where the doping is used to increase the PCM resistance.^[31,36] We note that all previous studies from literature reporting reset energy used pulses wider than 3 ns, thus this has not been examined in the range of fast pulse widths pursued in this work.

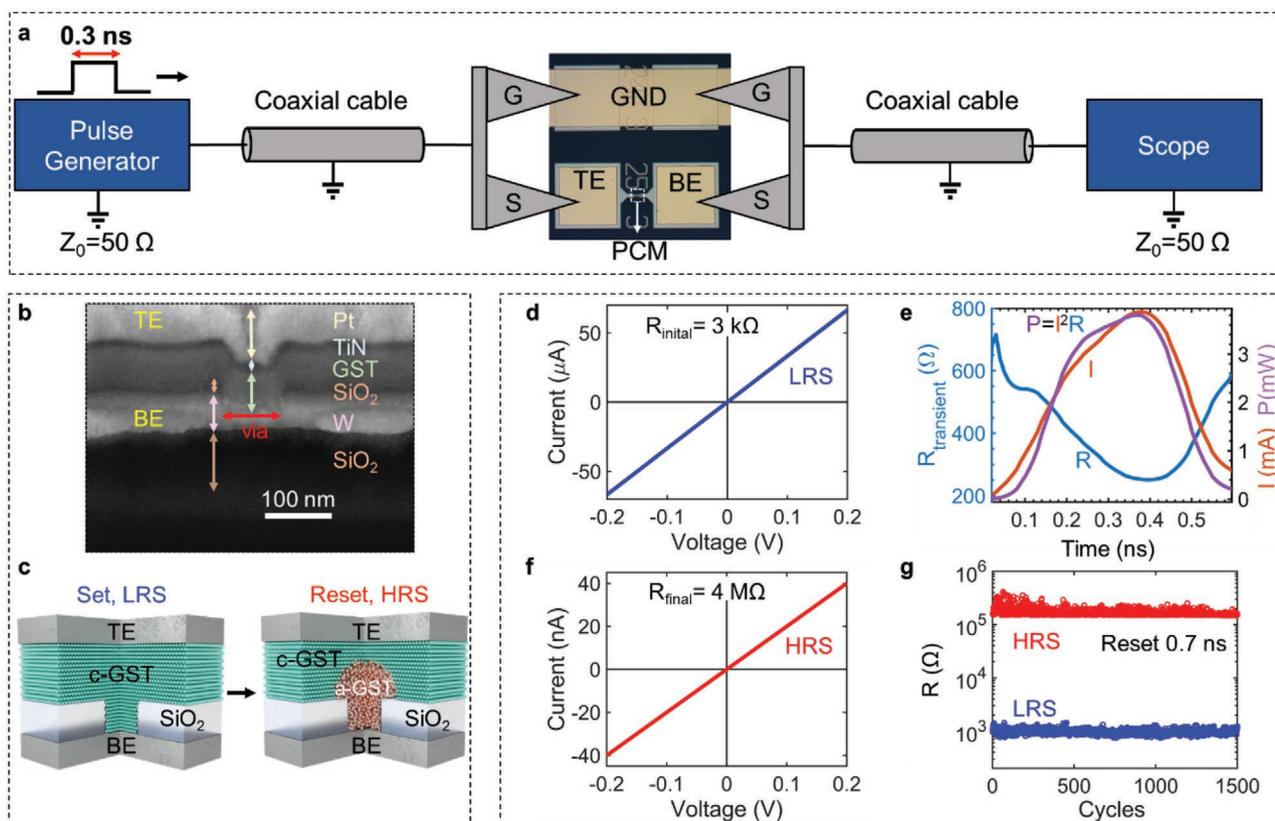


Figure 1. High-speed measurement setup and PCM device. a) Schematic of the measurement setup. A pulse generator (PG) with output resistance of 50Ω is connected by coaxial cable to an RF probe with the signal (S) to the top electrode (TE) pad of the PCM device and ground (G) to a local ground plane, shown in the optical image of the device. A fast scope is connected by RF probes (S) to the bottom electrode (BE) of the device. The RF probes are needed to minimize parasitic capacitance, and 50Ω impedance matching suppresses signal reflections. b) Cross section SEM image of a representative device. The phase change material is $\text{Ge}_2\text{Sb}_2\text{Te}_5$ (GST), sandwiched between Pt/TiN TE and W BE on a SiO_2/Si substrate. Sputtered SiO_2 serves as the insulator and the confined via hole diameter is marked by a horizontal red arrow. c) Schematic cartoon illustrating the PCM device in its LRS or set state (crystalline, left), and the HRS, reset state, where the confined volume is amorphous (right). The following three panels represent the measurements by their chronological order: d) DC I - V sweep to read the initial resistance value in the set state, e) reset pulse transient waveform, showing the evolution of the resistance ($V_{\text{PCM}}(t)/I_{\text{PCM}}(t)$), current ($I(t)$), and power ($I(t) \times V_{\text{PCM}}(t)$) during the pulse. The measured energy reported here is obtained by integrating $P(t)$ over time. The final resistance is read by another DC I - V sweep as shown in (f). Successful reset process results in HRS. g) Endurance measurement of the first 1500 cycles with sub-ns reset pulses. Details of the endurance program are found in Section 4, and comparison with longer reset PW is shown in Figure S2, Supporting Information.

Lines in Figure 2a,b represent calculated reset energy for a simple model of a GST sphere surrounded by SiO_2 (see Section 4 and Figure S3, Supporting Information). The model roughly captures reset energy scaling with device dimensions, but more importantly it shows the energy can be reduced by orders of magnitude through a reduction of the PW. Energy dependence on PW is more significant for smaller devices because at narrower PW the heat has less time to diffuse outward and is better confined in the active PCM volume. For instance, the thermal diffusivity of SiO_2 is $\alpha_{\text{ox}} \approx 0.01 \text{ cm}^2 \text{ s}^{-1}$, so the heat propagates a diffusion length $L_{\text{th}} \approx 1 \text{ nm}$ in 1 ps , $\approx 10 \text{ nm}$ in 0.1 ns , and $\approx 300 \text{ nm}$ in 100 ns , where $L_{\text{th}} = (\alpha t)^{1/2}$. Thus, for any PW longer than even a fraction of a nanosecond, the thermal diffusion length is greater than the device diameter, suggesting that a larger volume (e.g., dielectric) is heated outside the PCM cell than inside it.

It should also be noted that the shortest pulse to induce reset phase transition (by melt-quench) for a given cell size can be evaluated based on the speed of sound in the material

(vs $\approx 1 \text{ nm ps}^{-1}$). In other words, based on this speed limit, the shortest reset pulse for a $\approx 50 \text{ nm}$ PCM cell would be $\approx 50 \text{ ps}$. The adiabatic limit (solid black lines in Figure 2) is the energy to heat and melt the GST volume (including its latent heat of melting), without heating up its surrounding, that is, assuming the GST is perfectly thermally isolated. For reference, the black dash-dot line (dashed in Figure 2c) represents the energy required to charge a pair of interconnect lines in order to program a single bit in a $1 \text{ k} \times 1 \text{ k}$ crossbar array.^[15]

The measured reset energies of our PCM devices are also benchmarked against alternative emerging non-volatile memory (NVM) devices, namely spin torque transfer magnetic random access memory (RAM), conductive bridge RAM, and resistive RAM in Figure 2c.^[15] Our results show that reset energies within the same range of the PCM contacted by CNT (diameter $\approx 1\text{--}2 \text{ nm}$) can be achieved in devices with more than $100\times$ larger contact area (via diameter $\approx 50 \text{ nm}$) by reducing the PW. For example, reset energy of $\approx 0.2 \text{ pJ}$ at PW of 0.3 ns

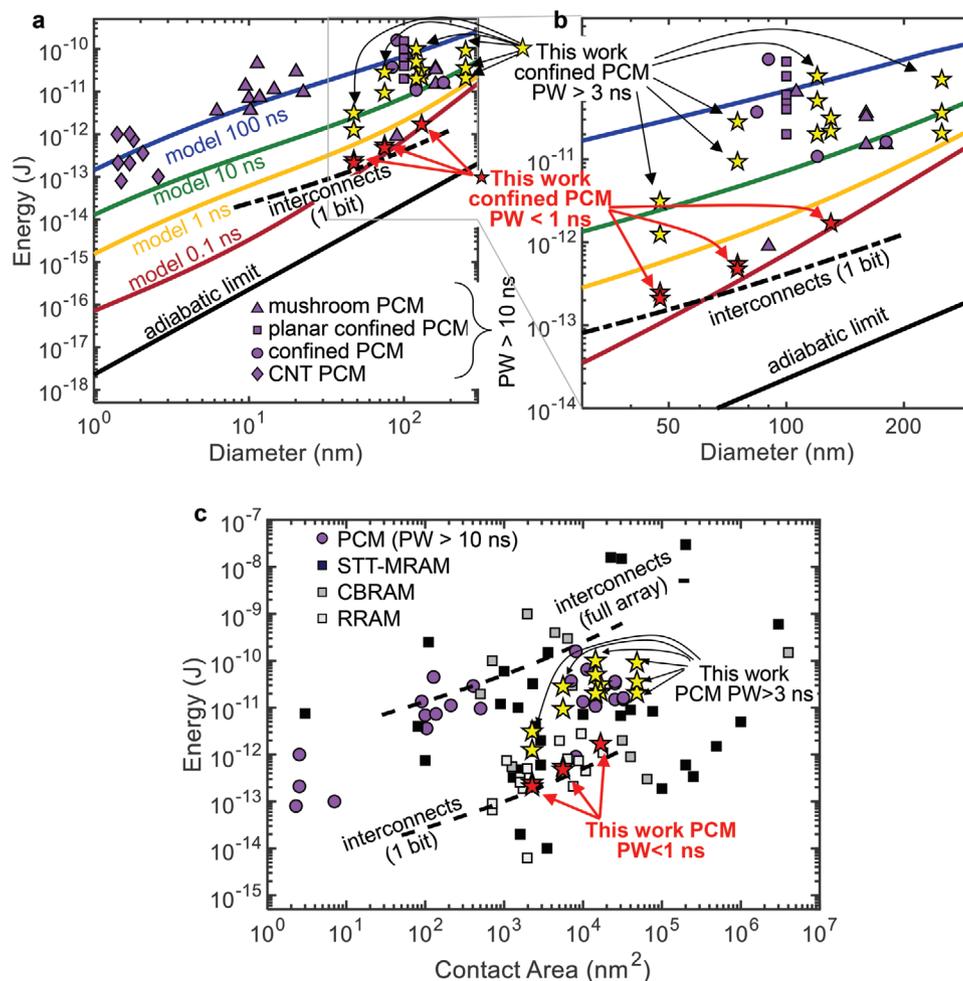


Figure 2. Benchmarking memory programming energy. a,b) PCM reset energy versus effective (confinement) diameter. Markers represent experimental data, lines represent calculations. All literature data of PCM reset energy consumption (purple markers) is for pulse width (PW) longer than 10 ns, whereas in this work we report on sub-ns reset programming (red stars) in addition to few ns to tens of ns pulses (yellow stars). The model lines represent calculated heating energy for GST sphere at T_M surrounded by SiO₂ with varying pulse width. The adiabatic limit assumes perfectly isolated PCM, and the interconnects energy (dashed line) is the energy to charge only the metal lines for a single bit in a $1\text{ k} \times 1\text{ k}$ crossbar array. c) Programming energy of different emerging non-volatile memory devices, including PCM (same as in (a)), spin-torque transfer magnetic RAM (STT-MRAM), conductive-bridge RAM (CBRAM), and resistive RAM (RRAM). Lower dashed line (interconnects) is same as in (a), and the upper one is for a full $1\text{ k} \times 1\text{ k}$ array.

was measured for ≈ 50 nm via, and ≈ 0.55 pJ at PW of 0.3 ns was measured for ≈ 75 nm via. The comparison to other NVM devices in Figure 2c shows that the PCM energy achieved with sub-ns PW is comparable to energy consumption in alternative technologies and approaches the interconnect charging energy, a practical lower-limit target.

Table 1 summarizes key results from the literature of reset energy consumption as well as sub-ns switching. Although several studies have shown sub-ns reset, the switching energy was not probed,^[24,26,28,37–39] due to the challenge in measuring the current and voltage dynamics at such short time-scale. Similarly, other studies have demonstrated reduction in reset energy by size scaling,^[22,29,33] and improved thermal barriers,^[17] but have not measured the switching energy for sub-ns pulses. Our results show that the reset energy is reduced as PW is decreased, and it should be emphasized that this reduction is in addition to other improvements in

the energy consumption such as size scaling and thermal barrier engineering. Therefore the projected energy consumption for a PCM cell with 10 nm feature size at sub-ns reset pulse would be below 10 fJ.

The improved PCM reset energy is lower than typical set energy consumption. Although the set power is lower than reset, the relatively slow crystallization process, which typically requires more than 50 ns in GST makes the overall set energy larger, for example, more than few pJ.^[15,40] Hence in a set-reset cycle, the fraction of the energy of the set process starts to dominate. Our sub-ns method is not valid for the set process in conventional PCM devices. This result puts more pressure on finding materials which crystallize on very fast time scales. Since the power consumption of set is roughly an order of magnitude lower than reset,^[15,41] materials which crystallize at sub-5 ns time scale are needed to obtain equal energies for the set and reset process.

Table 1. Comparison of state-of-the-art PCM reset energy and pulse width measurements.

Device	PW [ns]	Energy [pJ]	Energy density [fJ nm ⁻²]	Size [nm]	Transient waveform	Comments and ref.
GST mushroom	100	15	0.46	180	–	TaOx thermal barrier. ^[17]
GST in CNT crossbar	50	1	345	1.7	–	[33]
GST in CNT nano-gap	20	0.1	14	2.6	–	[22]
GST in CNT crossbar	10	0.2	70	1.7	–	[33]
GeTe mushroom	2.5 ^{a)}	7.5	2	60	Shown	[25]
SST mushroom	0.7	–	–	190	Shown for V	Study focused at reducing set time. ^[38]
GST confined	0.5	–	–	30	Shown for V	Study focused at reducing set time. ^[24]
GST crossbar	0.4	–	–	–	–	[28]
GST confined	0.4	–	–	20	–	[37]
GST confined	0.2	–	–	1000	–	[26]
GST confined	0.8	3.45	0.35	100	Shown	This work
GST confined	0.4	0.2	0.08	50	Shown	This work

^{a)}Overall PW: rise+plateau+fall = 2 + 1 + 2 ns

2.3. Phase Change Memory Reset Energy Limits and Power-Speed Trade-Off

PCM reset energy measurements show a clear benefit for PW reduction down to 0.3 ns, for devices with via diameter of 50–200 nm. Next, we turn to explore in more details the power-speed trade-off and energy limits of the reset process. **Figure 3** shows the final resistance following a reset pulse versus a) applied power and b) current with varying PWs for a 75 nm via device. Successful reset is defined here for $R_{\text{final}} > 100 \text{ k}\Omega$, that is, a change of nearly two orders of magnitude in resistance. Lower (higher) power and energy are required for smaller (larger) change in resistance, as shown in Figure S4, Supporting Information. We observe that for PWs from 40 ns down to 6 ns, there is no appreciable difference in the reset power.^[42] This is explained by the fact that the GST volume required to reset is already melted at <5 ns, and applying wider pulses mostly heats up the surrounding of the cell. It follows that for reset pulses longer than the thermal time constant of the GST region, reducing the PW will reduce the reset energy by

the same factor because the power remains unchanged. Below PW of $\approx 2 \text{ ns}$ the reset power increases with the reduction of the PW. Amorphization requires $\approx 3\times$ to $4\times$ higher power for 0.3 ns pulses compared with 6 ns (or longer). Although more power is needed, in terms of energy the 0.3 ns PW is preferred ($\approx 5\times$ lower reset energy). Measured and simulated power-energy curves are shown in Figure S5, Supporting Information. It should be noted that larger power, which is needed in the sub-ns regime, might degrade the endurance.^[43]

To better understand the relation between reset power and PW, we simulated the confined cells using FEM (see more details in Section 4 and Figure S6, Supporting Information) for different via diameters and varying PW. **Figure 4a,b** show the measured and simulated reset energy versus PW for varying via sizes. Reset energy scales with via size as expected. Similarly, reset energy scales with PW down to $\approx 1\text{--}2 \text{ ns}$, and the reset power remains constant in that range, as shown in Figure 4c,d. Below a critical PW the power increases. This critical time is shorter for smaller devices because it is proportional to the thermal time constant of the confined GST volume τ_{th} (where τ_{th} scales with device size).

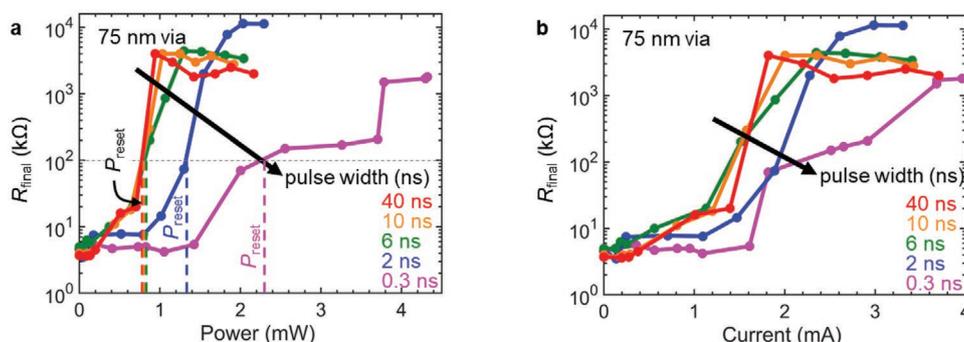


Figure 3. PCM reset power and current with varying pulse width. a) Final read resistance versus maximum measured power (during reset pulse) with varying pulse width (PW) in the range 0.3–40 ns for confined PCM with $\approx 75 \text{ nm}$ via. The reset power remains unchanged for pulses longer than $\approx 5 \text{ ns}$ and starts to increase as the PW is reduced below $\approx 3\text{--}4 \text{ ns}$. The reset powers (defined here for $R_{\text{final}} \approx 100 \text{ k}\Omega$) are indicated by vertical dashed lines. b) Final read resistance, same as (a) versus current. The initial resistance is in the range of 3–5 $\text{k}\Omega$. Rise and fall times are set to 70 ps.

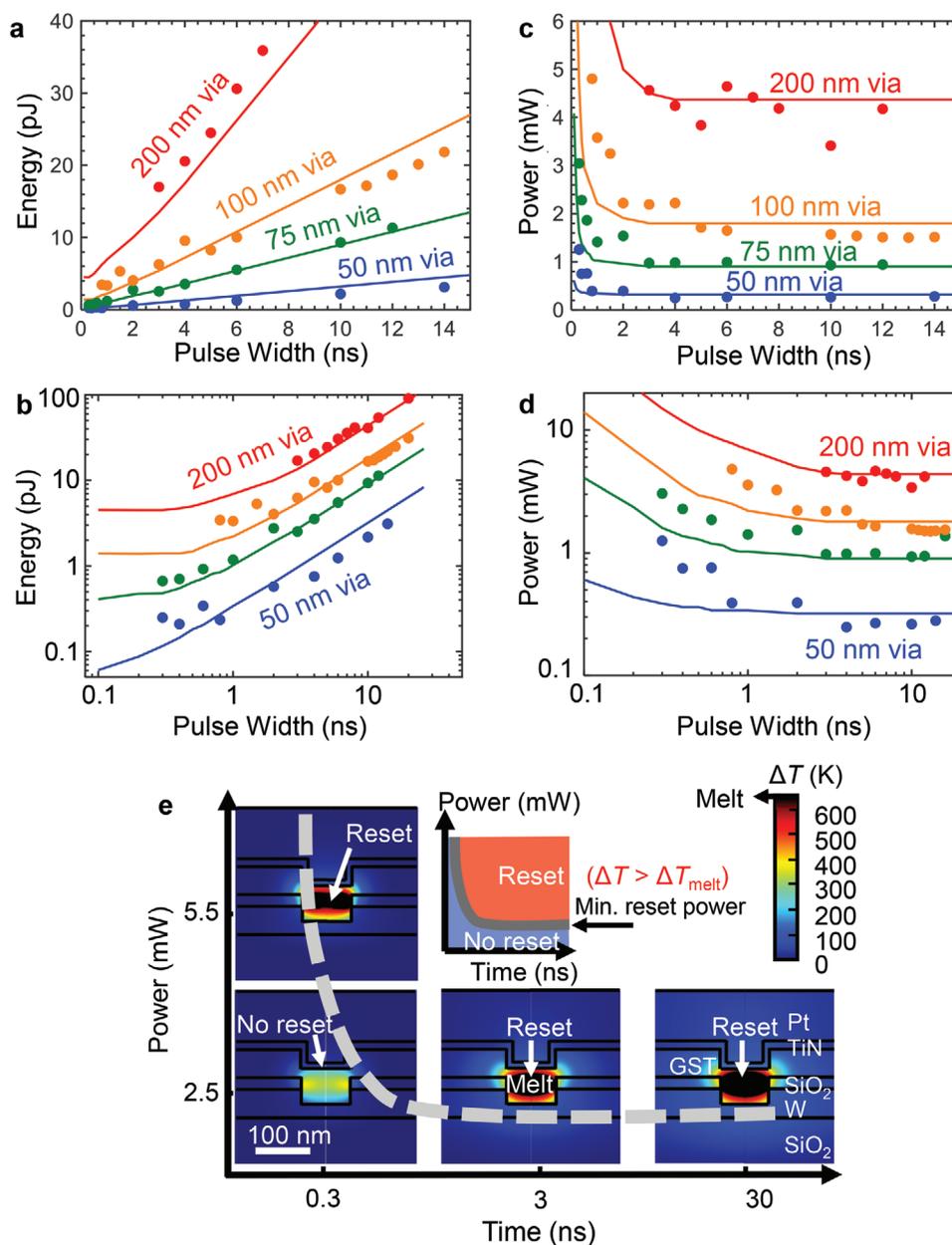


Figure 4. Switching energy and power-speed trade-off. a) Linear and b) log-log measured (markers) and simulated (lines) reset energy versus pulse width (PW) for devices with different via size (nominal =50, 75, 100, 200 nm). c) Linear and d) log-log measured (markers) and simulated (lines) reset power versus PW for devices with different via size (nominal 50, 75, 100, 200 nm). e) Maps of the temperature rise (ΔT) obtained by finite element electro-thermal simulations of the PCM devices used in this work, arranged on power versus time (log-scale) with a schematic description of the dependence of minimum switching power versus time (or pulse width). In the temperature maps, black corresponds to melted volume ($\Delta T > 600$ K). The trend shows that the minimum reset power is nearly constant for PW longer than few ns and starts to increase as the PW is reduced below ≈ 1 –2 ns (roughly the thermal transient of the confined GST).

This critical power-time point is predicted by our simulations and can be explained by the device temperature distribution (Figure 4e). Applying the minimal power for reset, the GST volume is melted within the thermal time constant of the confined GST, for example, ≈ 2 –3 ns for ≈ 100 nm via. For PW $> \tau_{th}$ (e.g., 30 ns) the only difference in the temperature map (compared with short PW $\approx \tau_{th}$) is further heating the surrounding of the cell (see Figure S7, Supporting Information). However, applying the same power with PW $< \tau_{th}$ (e.g., 0.3 ns) the GST is

not melted, and reset is not achieved. At such short PW, larger power is needed to reach the melting temperature (T_m) and achieve reset.

2.4. Phase Change Memory Thermal Dynamics

The thermal time constant is a critical property of the device that determines the optimal programming PW conditions

as outlined above. Besides simulations, an experimental approach is needed to evaluate the thermal dynamics of PCM. Previous work has shown that the resistivity of the GST in the liquid phase is smaller than in the crystalline phase (including high-temperature hexagonal GST).^[44] Thus, probing the transient resistance R_{trans} while the programming pulse is applied (Figure 1e) allows us to evaluate the thermal dynamics. During pulse application, R_{trans} starts to decrease as the GST reaches $\approx \Delta T_m$ and starts to melt. R_{trans} is further reduced as a larger volume of the GST is melted. Performing such measurements with sub-100 ps temporal resolution allows us to characterize the reset dynamics for sub-ns pulses.

Transient measurements with via sizes of 100, 75, and 50 nm are shown in Figure 5. The transient voltage applied by the pulse generator (V_{PG}) and the voltage measured by the scope (V_{scope}) are shown in Figure 5a for a representative 75 nm via device. The transient resistance is extracted as $R_{\text{trans}} = V_{\text{PCM}}(t)/I(t)$ and plotted in Figure 5b with the current and power versus time. R_{trans} scales inversely with via size (area), in agreement with a uniform (non-filamentary) melt-quench mechanism of PCM. The minimum measured values of R_{trans} are $\approx 440 \Omega$, $\approx 200 \Omega$, and 130Ω , for 50 nm, 75 nm, and 100 nm vias, respectively (Figure S8, Supporting Information). Normalized R_{trans} for different via sizes are depicted in Figure 5c, showing that the thermal transient is shorter for smaller devices. Previous work,^[26,28] suggested non-thermal amorphization for sub-ns reset, but the transient resistance waveforms that we uncover here combined with electro-thermal simulations can explain the thermally induced melting and reset.

We point out that as device dimensions are reduced, the interfaces and contacts become more dominant both electrically and thermally.^[45,46] Nonetheless, scaling down device dimensions is expected to effectively reduce the thermal time constant because the thermal capacitance keeps decreasing with PCM size. Finally, we note that for a given cell size, the final resistance is larger for lower R_{trans} (during pulse application), since larger melt volume during the pulse (low R_{trans}) results in larger amorphous volume after the pulse (larger R_{final}).

3. Conclusion

We have reported nanosecond and sub-nanosecond pulsing and probing in confined PCM cells, showing record-low reset switching energy density (e.g., less than $0.1 \text{ nJ } \mu\text{m}^{-2}$ at $\text{PW} = 0.3 \text{ ns}$). We show that the reset power remains almost constant for pulses wider than the GST thermal time constant τ_{th} . Extending the PW longer than τ_{th} ($\approx 1\text{--}2 \text{ ns}$) causes wasted energy by heating the dielectric surrounding the GST volume. For PW below τ_{th} , higher power is needed to achieve reset, yet the overall reset energy is reduced within the shortest PW probed in our measurements (0.3 ns). Hence, constrained power systems should work with reset programming pulses on the order of $\approx \tau_{\text{th}}$, whereas high-performance computation systems could prefer lower write latencies at the expense of higher power consumption. Constrained energy systems could work with PW shorter than τ_{th} (sub-ns) if they can tolerate larger power consumption, that is, larger instantaneous current or voltage. Finally, we also showed that the thermal dynamics of PCM can be probed by measuring the transient resistance of the device during pulse application. Our results shed new light on the energy limits of PCM operation and provide important guidelines for energy-efficient programming. Importantly, in the sub-ns regime, PCM reset energy approaches the charging energy of the interconnect lines in a crossbar array, a practical lower-limit benchmark for NVM and neuromorphic devices.

4. Experimental Section

Device Fabrication: The confined via-hole PCM was fabricated on a Si substrate, with 100 nm thermally grown SiO_2 . The process was started by defining the bottom electrode (BE). Tungsten (W) was evaporated, patterned with e-beam lithography, and dry-etched with reactive-ion etch. Next, SiO_2 was deposited using PECVD and patterned the via-hole using e-beam lithography. A 30-nm-thick GST layer capped in situ with 20 nm TiN was DC sputtered and lifted-off. The GST/TiN layer was deposited in situ after the Ar sputter cleaning to prevent any native oxide formation. After GST/TiN lift-off, an additional 20/40 nm sputtered TiN/Pt was patterned and lifted-off to form the probe pads and top electrode. Finally, a second metallization of 20/200 nm Ti/Au was e-beam evaporated and lifted-off, patterned with direct laser writer photolithography, to form the GS transmission line (TL).

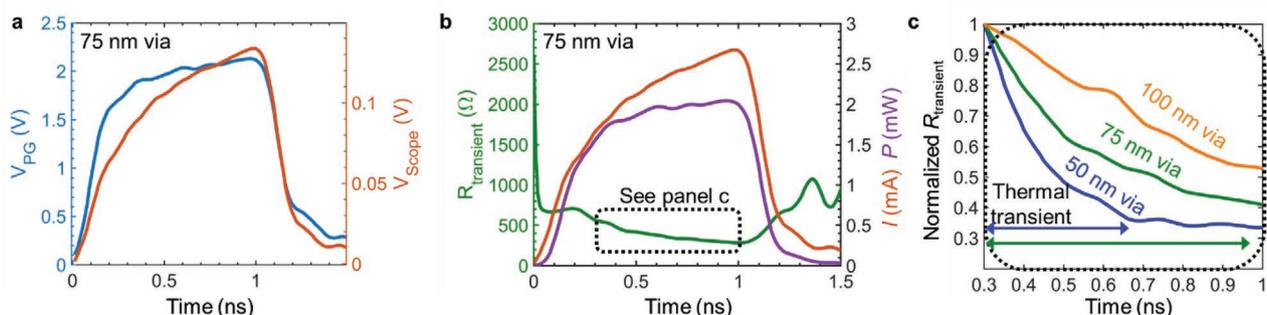


Figure 5. Dynamics of PCM reset. a) Measured transient waveform of the voltage applied by the pulse generator (V_{PG} , blue) and the voltage probed by the scope (orange) during a 1 ns pulse for 75 nm via device. b) Transient current ($I(t) = V_{\text{scope}}(t)/50 \Omega$), power ($P(t) = V_{\text{PCM}}(t) \cdot I(t)$), and resistance ($R_{\text{transient}} = V_{\text{PCM}}(t)/I(t)$) during the 1 ns waveform. The voltage across the PCM is obtained by $V_{\text{PCM}} = V_{\text{PG}} - V_{\text{scope}} - I(t) \times (Z_0 + R_S)$. c) Normalized transient resistance versus time for devices with different via size showing the thermal transient of the GST confinement.

Electrical Measurements: The experimental setup (Figure 1a) consisted of a fast pulse generator (PG) Active Technologies Pulse Rider PG-1072 connected in series with the PCM device and a Keysight Infiniium DSOS804A high-speed oscilloscope. The datasheet of the scope (<https://www.keysight.com/us/en/assets/7018-04261/data-sheets/5991-3904.pdf>) specifies typical rise/fall time of: 10/90%: 53.8 ps, 20/80%: 33.8 ps, and a temporal resolution of 3.15 ps. To support sub-nanosecond pulses, ground-signal (GS) transmission lines with characteristic impedance $Z_0 = 50 \Omega$ were used to connect the device with the measurement equipment. Transmission lines reduced pulse broadening, signal reflections, and possible ringing due to parasitic capacitances. The output impedance of the PG, as well as, the input impedance of the scope were also set to 50Ω . GGB GS RF probes were used in this setup, connected to the measurement equipment with 3.5 mm SMA cables.

Measurements were performed as follows. First, for each set of via sizes a “thru” configuration was measured, namely devices without GST were probed to obtain the transient waveform transmitted from the PG to the scope while applied to the cables, RF probes, device leads, without the PCM. Next, devices were programmed to LRS ($\approx 1 \text{ k}\Omega$ for 200-nm-via devices and $\approx 10 \text{ k}\Omega$ for the 50-nm-via). Set pulses were 100 μs long with varying amplitude in the range $\approx 1.4\text{--}2 \text{ V}$ depending on the via size. The resistance of the cell was read with a DC bias voltage of 0.2 V.

Then, reset pulses with varying amplitude (0.8–5 V) and with varying pulse width (0.3–40 ns) were applied. Each reset pulse was followed by a read operation with a DC voltage of 0.2 V and a set pulse to return to the initial low resistance. The current through the PCM ($I(t)$) was calculated from the measured voltage at the scope divided by its input impedance ($Z_0 = 50 \Omega$).

The endurance test was performed using Pulse Rider PG-1072 as pulse generator source and Keysight CX3322A Device Current Waveform Analyzer as current probe to read the resistance. Reset pulse widths were 0.7, 5, 10, and 40 ns with initial voltage amplitudes ranging from 6 V (for 0.7 ns PW) and between 3.6 and 4.8 V for the other PWs. Set PW was 5 μs with voltage amplitude of 2 V. The resistance was read by applying a 500 μs 0.2 V pulse and measuring the current using the CX3322A.

Focus-Ion Beam-Scanning Electron Microscopy: Cross-sectional images of the devices were made using focus-ion beam (FIB) SEM in a Helios NanoLab G3 series DualBeam. The top-surface of the device was protected with 300-nm Pt sputtered in situ. A cross-section was milled using a 30 kV and 40 pA Ga beam. High resolution SEM micrographs were acquired at 5 kV and 10 nA beam with a 52-tilt angle using through lens detector in topographical mode and in-column detector, which provided compositional contrast.

Simplified (Spherical) Reset Energy Model: The reset energy within the model is calculated as follows. A GST sphere was assumed having diameter D embedded in SiO_2 matrix. First, the adiabatic limit was calculated for which the GST sphere was perfectly isolated. The reset energy in this case is

$$E_{\text{ad}} = (C_s \times \Delta T_m + H) \cdot \text{Vol} \quad (3)$$

where C_s is the heat capacity of the GST, ΔT_m is the temperature rise required to reach the melting temperature, H is the latent heat of melting, and Vol is the GST volume. The other heating energies (with pulse width dependence) were obtained by first calculating the temperature rise distribution $\Delta T(r)$ for a given time t (corresponding to the pulse width) while the GST sphere is at $\Delta T = \Delta T_m$, followed by integrating $C_s \cdot \Delta T(r)$ over the entire space. A thermal boundary resistance (TBR) was included by adding its equivalent thermal resistance to the calculation of the temperature distribution. A relatively low value of the TBR was used $\approx 1 \text{ m}^2 \text{ K GW}^{-1}$, compared with measured TBR values for GST with its interfaces (SiO_2 , TiN, W) of $\approx 20\text{--}30 \text{ m}^2 \text{ K GW}^{-1}$ for two reasons: 1) The relevant TBR for the model was at high-temperature (near the melting temperature), whereas the measured values were around room temperature, and 2) the simplified model assumed GST embedded in SiO_2 , whereas real devices were contacted by metal electrodes which

had thermal conductivity more than 10x larger compared with insulators such as SiO_2 . Temperature maps of the model for varying PW are shown in Figure S3, Supporting Information.

Finite Element Method Electro-Thermal Simulations: Electro-thermal FEM simulations were performed in COMSOL, using a 2D axisymmetric finite element geometry. The bottom boundary of the Si substrate was held at ambient temperature, and the rest of the boundaries were insulating (adiabatic boundary condition). The BE was electrically grounded, and a voltage was applied to the TE. Device dimensions were taken from the FIB cross-sectional images (see Figure S1 and Table S1, Supporting Information). The model included electrothermal properties of the materials such as the temperature coefficient of resistance (TCR), the thermal conductivity (k_{th}), the heat capacity, and the TBR at the interfaces. Some of the key material parameters are outlined in Table S2, Supporting Information. The GST electrical conductivity was taken from the melt value reported in ref. [47], the electrical contact resistance was determined by an effective thickness $t_{\text{eff}} = 10 \text{ nm}$ of the GST as defined in ref. [19]. The thermal properties of the PCM were taken from the upper boundary in ref. [48] (because of the high temperature during reset). The TBR was chosen from the range of reported values in literature^[13,49] as a fitting parameter. More information on the distribution of voltage, current, and power density is shown in Figure S5, Supporting Information.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

K.S. and N.W. contributed equally to this work. This work was supported in part by ISF grant # 1179/20, by Russel Berrie Nanotechnology Institute (RBNI) seeding grant (NEVET), and by member companies of the Stanford Non-volatile Memory Technology Research Initiative (NMTRI). Fabrication of the confined PCM cells was performed at the Stanford Nanofabrication Facility (SNF) and Stanford Nano Shared Facilities (SNSF). Fabrication of the transmission lines was carried out at the Technion Micro-Nano Fabrication & Printing Unit (MNF&PU). The authors would like to thank Dr. Larisa Popilevsky, RBNI FIB Lab. The authors would also like to thank Ido Kaminer, Dan Ritter, and Mario Lazna for careful reading of the manuscript. E.Y. thanks Kye Okabe for fruitful discussions.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Keywords

energy efficiency, non-volatile memory, phase change materials, sub-nanosecond probing, thermal management

Received: March 3, 2021

Revised: May 17, 2021

Published online: June 12, 2021

- [1] T. Kim, S. Lee, *IEEE Trans. Electron Devices* **2020**, *67*, 1394.
- [2] S. W. Fong, C. M. Neumann, H. P. Wong, *IEEE Trans. Electron Devices* **2017**, *64*, 4374.
- [3] G. W. Burr, R. M. Shelby, A. Sebastian, S. Kim, S. Kim, S. Sidler, K. Virwani, M. Ishii, P. Narayanan, A. Fumarola, L. L. Sanches, I. Boybat, M. L. Gallo, K. Moon, J. Woo, H. Hwang, Y. Leblebici, *Adv. Phys.: X* **2017**, *2*, 89.
- [4] E. Yalon, I. M. Datye, J. S. Moon, K. A. Son, K. Lee, E. Pop, *IEEE Electron Device Lett.* **2019**, *40*, 455.
- [5] N. Wainstein, G. Adam, E. Yalon, S. Kvatinsky, *Proc. IEEE* **2020**, *109*, 77.
- [6] D. Ielmini, H. S. P. Wong, *Nat. Electron.* **2018**, *1*, 333.
- [7] M. L. Gallo, A. Sebastian, R. Mathis, M. Manica, H. Giefers, T. Tuma, C. Bekas, A. Curioni, E. Eleftheriou, *Nat. Electron.* **2018**, *1*, 246.
- [8] A. Sebastian, M. L. Gallo, R. Khaddam-Aljameh, E. Eleftheriou, *Nat. Nanotechnol.* **2020**, *15*, 529.
- [9] G. De Sandre, L. Bettini, A. Pirola, L. Marmonier, M. Pasotti, M. Borghi, P. Mattavelli, P. Zuliani, L. Scotti, G. Mastracchio, F. Bedeschi, R. Gastaldi, R. Bez, *IEEE J. Solid-State Circuits* **2011**, *46*, 52.
- [10] A. Fazio, presented at IEEE Int. Electron Devices Meet., 24, San Francisco, CA, USA, December **2020**.
- [11] S. R. Ovshinsky, *Phys. Rev. Lett.* **1968**, *21*, 1450.
- [12] S. Raoux, F. Xiong, M. Wuttig, E. Pop, *MRS Bull.* **2014**, *39*, 703.
- [13] H. S. P. Wong, S. Raoux, S. Kim, J. Liang, J. P. Reifenberg, B. Rajendran, M. Asheghi, K. E. Goodson, *Proc. IEEE* **2010**, *98*, 2201.
- [14] F. Xiong, E. Yalon, A. Behnam, C. M. Neumann, K. L. Grosse, S. Deshmukh, E. Pop, presented at IEEE Int. Electron Devices Meet., 79, San Francisco, CA, USA, December **2016**, 79.
- [15] H.-S. P. Wong, C. Ahn, J. Cao, H.-Y. Chen, S. B. Eryilmaz, S. W. Fong, J. A. Incorvia, Z. Jiang, H. Li, C. Neumann, K. Okabe, S. Qin, J. Sohn, Y. Wu, S. Yu, X. Zhang, Stanford Memory Trends, <https://nano.stanford.edu/stanford-memory-trends> (accessed: June **2021**).
- [16] A. L. Lacaíta, A. Redaelli, *Microelectron. Eng.* **2013**, *109*, 351.
- [17] Y. Matsui, K. Kurotsuchi, O. Tonomura, T. Morikawa, M. Kinoshita, Y. Fujisaki, N. Matsuzaki, S. Hanzawa, M. Terao, N. Takaura, H. Moriya, T. Iwasaki, M. Moniwa, T. Koga, presented at IEEE Int. Electron Devices Meet., 2, San Francisco, CA, USA, December **2006**.
- [18] S. W. Fong, C. M. Neumann, E. Yalon, M. M. Rojo, E. Pop, H. S. P. Wong, *IEEE Trans. Electron Devices* **2017**, *64*, 4496.
- [19] C. M. Neumann, K. L. Okabe, E. Yalon, R. W. Grady, H.-S. P. Wong, E. Pop, *Appl. Phys. Lett.* **2019**, *114*, 082103.
- [20] C. Ahn, S. W. Fong, Y. Kim, S. Lee, A. Sood, C. M. Neumann, M. Asheghi, K. E. Goodson, E. Pop, H. S. P. Wong, *Nano Lett.* **2015**, *15*, 6809.
- [21] D. H. Im, J. I. Lee, S. L. Cho, H. G. An, D. H. Kim, I. S. Kim, H. Park, D. H. Ahn, H. Horii, S. O. Park, U. I. Chung, J. T. Moon, presented at IEEE Int. Electron Devices Meet., San Francisco, CA, USA, December **2008**.
- [22] F. Xiong, A. D. Liao, D. Estrada, E. Pop, *Science* **2011**, *332*, 568.
- [23] E. Pop, *Nano Res.* **2010**, *3*, 147.
- [24] D. Loke, T. H. Lee, W. J. Wang, L. P. Shi, R. Zhao, Y. C. Yeo, T. C. Chong, S. R. Elliott, *Science* **2012**, *336*, 1566.
- [25] G. Bruns, P. Merklebach, C. Schlockermann, M. Salinga, M. Wuttig, T. D. Happ, J. B. Philipp, M. Kund, *Appl. Phys. Lett.* **2009**, *95*, 043108.
- [26] D. Q. Huang, X. S. Miao, Z. Li, J. J. Sheng, J. J. Sun, J. H. Peng, J. H. Wang, Y. Chen, X. M. Long, *Appl. Phys. Lett.* **2011**, *98*, 242106.
- [27] D. Loke, J. M. Skelton, W. J. Wang, T. H. Lee, R. Zhao, T. C. Chong, S. R. Elliott, *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 13272.
- [28] P. Wang, C. Ju, W. Chen, D. Q. Huang, X. W. Guan, Z. Li, X. M. Cheng, X. S. Miao, *Appl. Phys. Lett.* **2013**, *102*, 112108.
- [29] F. Xiong, M. H. Bae, Y. Dai, A. D. Liao, A. Behnam, E. A. Carrion, S. Hong, D. Ielmini, E. Pop, *Nano Lett.* **2013**, *13*, 464.
- [30] H. R. Oh, B. H. Cho, W. Y. Cho, S. Kang, B. G. Choi, H. J. Kim, K. S. Kim, D. E. Kim, C. K. Kwak, H. G. Byun, G. T. Jeong, H. S. Jeong, K. Kim, *IEEE J. Solid-State Circuits* **2006**, *41*, 122.
- [31] S. Kim, M. Ishii, S. Lewis, T. Perri, M. Brightsky, W. Kim, R. Jordan, G. W. Burr, N. Sosa, A. Ray, J. P. Han, C. Miller, K. Hosokawa, C. Lam, presented at IEEE Int. Electron Devices Meet., 17, Washington, DC, USA, December **2015**.
- [32] R. Annunziata, P. Zuliani, M. Borghi, G. De Sandre, L. Scotti, C. Prelini, M. Tosi, I. Tortorelli, F. Pellizzer, presented at IEEE Int. Electron Devices Meet., 97, Baltimore, MD, USA, December **2009**.
- [33] J. Liang, R. G. D. Jeyasingh, H. Y. Chen, H. S. P. Wong, *IEEE Trans. Electron Devices* **2012**, *59*, 1155.
- [34] E. J. Evans, J. H. Helbers, S. R. Ovshinsky, *J. Non-Cryst. Solids* **1970**, *2*, 334.
- [35] S. H. Lee, H. C. Park, M. S. Kim, H. W. Kim, M. R. Choi, H. G. Lee, J. W. Seo, S. C. Kim, S. G. Kim, S. B. Hong, S. Y. Lee, J. U. Lee, Y. S. Kim, K. S. Kim, J. I. Kim, M. Y. Lee, H. S. Shin, S. J. Chae, J. H. Song, H. S. Yoon, J. M. Oh, S. K. Min, H. M. Lee, K. R. Hong, J. T. Cheong, S. N. Park, J. C. Ku, H. S. Shin, Y. S. Sohn, S. K. Park, et al., *Tech. Dig. - Int. Electron Devices Meet.* **2011**, *2*, 47.
- [36] H. Y. Cheng, J. Y. Wu, R. Cheek, S. Raoux, M. Brightsky, D. Garbin, S. Kim, T. H. Hsu, Y. Zhu, E. K. Lai, E. Joseph, A. Schrott, S. C. Lai, A. Ray, H. L. Lung, C. Lam, presented at IEEE Int. Electron Devices Meet., San Francisco, CA, USA, December **2012**.
- [37] W. J. Wang, L. P. Shi, R. Zhao, K. G. Lim, H. K. Lee, T. C. Chong, Y. H. Wu, *Appl. Phys. Lett.* **2008**, *93*, 043121.
- [38] F. Rao, K. Ding, Y. Zhou, Y. Zheng, M. Xia, S. Lv, Z. Song, S. Feng, I. Ronneberger, R. Mazzarello, W. Zhang, E. Ma, *Science* **2017**, *358*, 1423.
- [39] D. K. Loke, J. M. Skelton, T. H. Lee, R. Zhao, T. C. Chong, S. R. Elliott, *ACS Appl. Mater. Interfaces* **2018**, *10*, 41855.
- [40] J. Wang, X. Dong, G. Sun, D. Niu, Y. Xie, presented at IEEE Int. Conf. Computer Design: VLSI Comp. and Processors, 175, Amherst, MA, September **2011**.
- [41] D. Krebs, S. Raoux, C. T. Rettner, G. W. Burr, R. M. Shelby, M. Salinga, C. M. Jefferson, M. Wuttig, *J. Appl. Phys.* **2009**, *106*, 054308.
- [42] E. Yalon, K. Okabe, C. M. Neumann, H. S. P. Wong, E. Pop, presented at IEEE Device Res. Conf. (DRC), Santa Barbara, CA, USA, June **2018**.
- [43] S. Lee, J. H. Jeong, T. S. Lee, W. M. Kim, B. K. Cheong, *IEEE Electron Device Lett.* **2009**, *30*, 472.
- [44] K. Cil, F. Dirisaglik, L. Adnane, M. Wennberg, A. King, A. Faraclas, M. B. Akbulut, Y. Zhu, C. Lam, A. Gokirmak, H. Silva, *IEEE Trans. Electron Devices* **2013**, *60*, 433.
- [45] J. P. Reifenberg, D. L. Kencke, K. E. Goodson, *IEEE Electron Device Lett.* **2008**, *29*, 1112.
- [46] D. L. Kencke, I. V. Karpov, B. G. Johnson, S. J. Lee, D. Kau, S. J. Hudgens, J. P. Reifenberg, S. D. Savransky, J. Zhang, M. D. Giles, G. Spadini, presented at IEEE Int. Electron Devices Meet., 323, Washington, DC, USA, December **2007**.
- [47] N. Wainstein, G. Ankonina, S. Kvatinsky, E. Yalon, *IEEE Trans. Electron Devices* **2020**, *67*, 5182.
- [48] E. Bozorg-Grayeli, J. P. Reifenberg, M. Asheghi, H.-S. P. Wong, K. E. Goodson, *Annu. Rev. Heat Transfer* **2013**, *16*, 397.
- [49] E. Yalon, S. Deshmukh, M. M. Rojo, F. Lian, C. M. Neumann, F. Xiong, E. Pop, *Sci. Rep.* **2017**, *7*, 15360.